

On Counting Cliques, Clique-covers and Independent sets in Random Graphs^{*}

Kashyap Dixit and Martin Fürer

Pennsylvania State University
111 IST Building, University Park 16801, USA
kashyap@cse.psu.com
furer@cse.psu.edu

Abstract. We study the problem of counting the number of *isomorphic* copies of a given *template* graph, say H , in the input *base* graph, say G . In general, it is believed that polynomial time algorithms that solve this problem exactly are unlikely to exist. So, a lot of work has gone into designing efficient *approximation schemes*, especially, when H is a perfect matching. In this work, we present efficient approximation schemes to count k -Cliques, k -Independent sets and k -Clique covers in random graphs.

We present *fully polynomial time randomized approximation schemes* (fpras) to count k -Cliques and k -Independent sets in a random graph on n vertices when k is at most $(1 + o(1)) \log n$, and k -Clique covers when k is a constant. The problem of counting k -cliques and k -independent sets was an open problem in [Frieze and McDiarmid, 1997]. In other words, we have a fpras to evaluate the first $(1 + o(1)) \log n$ terms of the *clique polynomial* and the *independent set polynomial* of a random graph. [Grimmett and McDiarmid, 1975] present a simple greedy algorithm that *detects* a clique (independent set) of size $(1 + o(1)) \log_2 n$ in $G \in \mathcal{G}(n, \frac{1}{2})$ with high probability. No algorithm is known to detect a clique or an independent set of larger size with non-vanishing probability. Furthermore, [Coja-Oghlan and Efthymiou, 2011] present some evidence that one cannot hope to easily improve a similar, almost 40 years old bound for sparse random graphs. Therefore, our results are unlikely to be easily improved.

We use a novel approach to obtain a recurrence corresponding to the variance of each estimator. Then we upper bound the variance using the corresponding recurrence. This leads us to obtain a polynomial upper bound on the critical ratio. As an aside, we also obtain an alternate derivation of the closed form expression for the k -th moment of a binomial random variable using our techniques. The previous derivation [Knoblauch (2008)] was based on the moment generating function of a binomial random variable.

Keywords: Random Sampling, Approximate Counting, Randomized Approximation Schemes for #P-complete problems.

^{*} supported in part by NSF Grant CCF-1320814

1 Introduction

Given a *base graph* G and a *template graph* H , the *subgraph isomorphism* problem is to decide whether an edge preserving injection ϕ between the vertices of H and G exists. That is, for every edge $\{u, v\}$ in H , $\{\phi(u), \phi(v)\}$ is an edge in G . Subgraph isomorphism is a generalization of several fundamental NP-complete problems, like Hamiltonian Path and Clique. The problem has applications in many areas, including cheminformatics [32], pattern discovery in databases [24], bioinformatics [27] and social networks [1].

Another widely studied related fundamental problem is that of counting the number of copies of H in G . In general, this problem is #P-complete (Valiant [33]). The class #P is defined as $\{f : \text{There exists a non-deterministic polynomial time Turing machine } M, \text{ such that on input } x, \text{ the computation tree of } M \text{ has exactly } f(x) \text{ accepting leaves}\}$. The problems complete in this class are computationally quite difficult, since an oracle access to #P complete problem would make it possible to solve any problem in the polynomial hierarchy in polynomial time (Toda [31]).

The k -Clique problem asks whether there exists a k -clique in the input graph G . A k -Clique is the complete graph on k vertices. The k -Clique problem has numerous applications, particularly in bioinformatics and social networks [27,1]. Counting k -cliques in a web-graph has applications in social network analysis. In particular, this gives an estimate of the number of closed communities in the web-graphs. Therefore, fast algorithms for counting k -cliques in web-graphs give an insight to the evolution of Internet.

The k -Clique cover problem asks for the existence of a perfect k -clique packing in G . More precisely, given base graph G with n vertices and template graph H that is n/k vertex disjoint and edge disjoint copies of k -cliques, does there exist an injective mapping from H to G . The decision problem k -Clique Cover, that is $\{(G, k) : \text{There exists a disjoint cover of } G \text{ by } k\text{-cliques}\}$ is NP-complete on general graphs with clique number 3 [21]. The k -Clique cover problem has applications in the *orgy problem* [7]: Given a group of people with affinities and aversion between them, is it possible to divide them into k members each, such that every person in each group is compatible with every other person in the group. Some of the scheduling problems can also be modeled as an orgy problem. We are given n jobs of length $\leq T$ seconds and n/k machines. Also, for each job j , we are given a list of *conflicting jobs* which can not be scheduled with j on the same machine. The problem is to schedule the jobs on the machines such that the total time to complete all the jobs is minimized.

The clique-polynomial [15] of a graph $G = (V, E)$ is given by $1 + \sum_{i=1}^{\omega(G)} c_i x^i$. Here, c_i denotes the number of i -Cliques in G , $\omega(G)$ denotes the size of largest clique in G . The independent-set polynomial [15] of a graph is defined analogously. In general, computing the clique-polynomial and the independent set polynomial of a graph G is #P-complete.

We consider template graphs which are vertex disjoint union of cliques. More specifically, we will be considering problems of counting cliques and clique covers. We note that our techniques can be extended to counting embeddings of template

graphs which are disjoint union of cliques of possibly different sizes. The counting version of the k -Clique problem is #P-complete in general. The counting version of the k -Clique cover problem is #P-complete even for $k = 2$ (Valiant ([33])), where H is a perfect matching.

Note that the counting versions of the aforementioned problems are extremely hard even for the simple cases. So, we try to come up with *fully polynomial time approximation schemes* (abbreviated as fpras) for these problems that work well for *almost all* graphs. More precisely, fpras must run in time $\text{poly}(n, \varepsilon^{-1})$ and return an answer within a relative error of $(1 \pm \varepsilon)$ with high probability (i.e., probability tending to 1 as $n \rightarrow \infty$) for graphs that are uniformly randomly sampled from $G \in \mathcal{G}(n, p)$. Here, $\mathcal{G}(n, p)$ denotes the class of graphs in which each edge occurs with probability p . Note that when $p = \frac{1}{2}$, each graph $G \in \mathcal{G}(n, p)$ is equiprobable. Another commonly studied model is $\mathbb{G}(n, m)$ where each graph with n vertices and m edges is assigned the same probability, which is $\binom{N}{m}^{-1}$, where $N = \binom{n}{2}$.

The theory of random graphs was initiated by Erdős and Rényi [9]. We work with the model $\mathcal{G}(n, p)$ where we are given a fixed set of n vertices and each of the $\binom{n}{2}$ edges is added with probability p .

Our analysis also provides an alternate derivation of the closed form of the k^{th} moment of a binomial random variable X sampled from $\text{Binomial}(n, p)$, which has been derived by Knoblach [23] using *moment generating function*. We derive the same results using simple binomial equalities that we obtain using the binomial theorem.

1.1 Our results

In this work, we present new results for k -Clique and k -Clique cover counting problems in random graphs. Our algorithm is based on the idea of Rasmussen's unbiased estimator for permanents [28]. It has been widely used in the context of subgraph isomorphism counting problems [29,11,12]. For counting k -cliques in the input random graph G , we embed a k -clique into G , doing so one vertex at a time chosen randomly. If the procedure succeeds, we compute the probability with which the clique is obtained in G and output its inverse. As shown in [12], this is an unbiased estimate of the number of cliques in G . We state the results below in Theorem 1. In this work, we generalize Rasmussen's approach [28] to efficiently count k -cliques and k -clique covers in random graphs. As a corollary, we also get a fpras for counting k -independent sets in random graphs. Note that [6] indicates that our bounds is extremely difficult to be improved.

Theorem 1. *Let H be a k -Clique, where $k = (1+o(1)) \log_{\frac{1}{p}} n$. Then, there exists a fpras for estimating the number of copies of H in $G \in \mathcal{G}(n, p)$ for constant p .*

Note that counting k -cliques in $\mathcal{G}(n, p)$ is equivalent to counting k -independent sets in $\mathcal{G}(n, 1-p)$. Since p is a constant in our case, we have a fpras for counting k -Independent sets of a random graph.

Theorem 2. *Let H be a k -independent set, where $k = (1 + o(1)) \log_{\frac{1}{p}} n$. Then, there exists a fpras for estimating the number of copies of H in $G \in \mathcal{G}(n, 1 - p)$ for constant p .*

For counting k -clique cover, we embed one clique at a time, until the whole graph is covered by k -cliques. The *key* observation here is that after embedding a clique, the residual base graph still remains random with edge probability p . We obtain the following theorem for counting k -clique covers.

Theorem 3. *Let H be a k -clique cover, where $k = O(1)$. Then, there exists a fpras for estimating the number of copies of H in $G \in \mathcal{G}(n, p)$ for constant p .*

Our estimators for counting cliques and clique-covers are given in Algorithm 1 and Algorithm 2 respectively in Section 4. As a side result, we obtain an alternate derivation of $\mathbf{E}[X^k]$ for a binomial random variable X , for all $k \geq 0$. We note that this has already been obtained in [23] using the moment generating function for binomial random variable.

Outline of the paper: In Section 2, we give some of the related work to set perspective for our work. To introduce our techniques to the reader, we give a new derivation for the closed form of k -th moment for binomial random variables using these techniques Section 3. We move on to describe estimators for counting k -cliques and k -clique covers in Section 4. We analyze these estimators for counting k -cliques and k -clique covers for random graphs in Section 5.1 and Section 5.2 respectively, which is the main contribution of this paper.

2 Related work

A lot of work has been done in finding and counting of cliques and independent sets in graphs. One of the earliest result in the theory of random graphs is about showing that the independence number and clique number of a random graph $G \in \mathcal{G}(n, \frac{1}{2})$ is about $2 \log_2 n$. Grimmett and McDiarmid [14] analyzed simple greedy algorithm constructs an inclusion-maximal independent set. They showed that it yields an independent set of size $(1 + o(1)) \log_2 n$. Coja-Oghlan and Efthymiou [6] show some evidence for why no better algorithm could be found over many years.

Luby and Vigoda [26] have shown a fully polynomial time scheme for counting independent sets in the graphs with maximum degree $\Delta \leq 4$, which was later improved by Weitz [34] to $\Delta \leq 5$. On the other hand, Dyer, Freize and Jerrum [8] have shown that no fpras exists for counting independent sets in graphs with $\Delta \geq 25$ unless $\text{NP}=\text{RP}$. They also show that the Markov Chain Monte Carlo technique is likely to fail if $\Delta \geq 6$. Chandrasekaran et.al. [4] have obtained fpras for higher degree graphs with large girths.

A major breakthrough in counting perfect matchings (2-clique covers) was a polynomial time algorithm for planar graphs due to Kasteleyn [22]. For a bipartite graph, it corresponds to calculating the permanent of a $\{0, 1\}$ matrix. In the seminal paper of Valiant [33], it has been shown to be $\#\text{P}$ -complete, even

though the decision version of this problem is in P. The noted work of Jerrum, Sinclair and Vigoda [18] presents a fpras for counting perfect matchings in bipartite graphs. The problem of existence and counting of covers in random graphs $G \in \mathcal{G}(n, p)$ was addressed in the seminal work of Johansson, Kahn and Vu [19]. They show that given a subgraph H , the number of H -covers in a random graph $G \in \mathcal{G}(n, p)$ is $e^{-O(n)}(n^{v-1}p^m)^{n/v}$ for large enough n with probability at least $1 - n^{-\Omega(1)}$. Here $v = |V(H)|$ and $m = |E(H)|$. Various approaches for getting an unbiased estimator with small variance have been explored for counting perfect matchings in other graphs. Some of these are determinant based approaches [13,20,5,25], Markov chain Monte Carlo (MCMC) algorithms [3,17,18,2] and search based on Rasmussen's techniques [28,29,11,12]. Chien [5] gives an efficient fpras for counting perfect matchings in random graphs. MCMC algorithms are polynomial time algorithms for all bipartite graphs. The estimators based on Rasmussen's approach (from [28]) have also been proved to work well in random graphs, where they lead to simple, polynomial time approximation schemes. In this work, we generalize Rasmussen's approach to efficiently count k -Cliques and k -Clique covers in random graphs. As a corollary, we also get a fpras for counting k -Independent sets in random graphs.

Rasmussen [29] has given a fpras for counting cliques and independent sets in random graphs. But it is unclear how to extend that algorithm for counting k -cliques [10] or k -Independent sets in random graphs. We note here that Fürer and Kasivaswanathan [12] have used similar techniques to get fpras for a large class of subgraph isomorphism problems. A fundamental constraint in their analysis was that the template subgraphs triangle-free. Thus, their analysis could not be extended directly to get fpras for k -clique, k -independent set and k -clique cover problems.

3 k^{th} moment of a binomial random variable

Consider the binomial random variable $X = \text{binomial}(n, p)$. We are interested in finding the k^{th} moment of X , i.e. we want to find $\mathbf{E}[X^k]$. In this section, we give the closed form expression for $\mathbf{E}[X^k]$. We evaluate using new equalities obtained from well known binomial theorem. Note that

$$\mathbf{E}[X^k] = \sum_{i=0}^n i^k \binom{n}{i} p^i (1-p)^{n-i}$$

We start with the most fundamental equality known as binomial theorem given below.

$$(1+x)^n = \sum_{i=0}^n \binom{n}{i} x^i \tag{1}$$

Suppose we differentiate (1) with respect to x and multiply by x subsequently, we get the following equation.

$$nx(1+x)^{n-1} = \sum_{i=0}^n i \binom{n}{i} x^i \quad (2)$$

Note that substituting $x = \frac{p}{1-p}$ in (2) and multiplying by $(1-p)^n$, we get $np = \sum_{i=0}^n i \binom{n}{i} p^i (1-p)^{n-i}$, which is the first moment of X . Suppose we differentiate (2) w.r.t. x again and multiply by x subsequently, we get

$$x(1+x)^{n-1}(n)_1 + x^2(1+x)^{n-2}(n)_2 = \sum_{i=0}^n i^2 \binom{n}{i} x^i \quad (3)$$

The term $(n)_i$ denotes the falling factorial $n \cdot (n-1) \cdot (n-2) \cdots (n-i+1) = \frac{n!}{(n-i)!}$. Again, substituting $x = \frac{p}{1-p}$ in (3) and multiplying $(1-p)^n$, we get $(n)_1 p + (n)_2 p^2 = \sum_{i=0}^n i^2 \binom{n}{i} p^i (1-p)^{n-i} = \mathbf{E}[X^2]$. The above calculations show an emerging pattern for higher moments, which Lemma 1 illustrates.

Lemma 1.

$$g(x, k) = \sum_{i=0}^n i^k \binom{n}{i} x^i = \sum_{j=1}^k \lambda_{k,j} x^j (1+x)^{n-j} (n)_j \quad (4)$$

Here $\lambda_{k,j}$ are the coefficients that depend on k and j but are independent of n . Here $0 \leq j \leq k$, $\lambda_{k,0} = \lambda_{k,k+1} = 0$.

Proof. We will prove the above lemma by induction. For $i = 1$, this is true as shown in (2). Suppose the lemma is true for $g(x, 1), g(x, 2), \dots, g(x, k)$. We prove that it holds for $g(x, k+1)$. Differentiating (4) w.r.t. x and subsequently multiplying with x gives

$$\begin{aligned} \sum_{i=0}^n i^{k+1} \binom{n}{i} x^i &= \sum_{j=1}^k \lambda_{k,j} (n)_j (jx^j (1+x)^{n-j} + (n-j)x^{j+1} (1+x)^{n-j-1}) \\ &= \sum_{j=1}^k \lambda_{k,j} j x^j (1+x)^{n-j} (n)_j + \sum_{j=1}^k \lambda_{k,j} x^{j+1} (1+x)^{n-j-1} (n-j) (n)_j \\ &= \sum_{j=1}^k \lambda_{k,j} j x^j (1+x)^{n-j} (n)_j + \sum_{j=1}^k \lambda_{k,j} x^{j+1} (1+x)^{n-j-1} (n)_{j+1} \\ &= \sum_{j=1}^{k+1} (j\lambda_{k,j} + \lambda_{k,j-1}) x^j (1+x)^{n-j} (n)_j \\ &= \sum_{j=1}^{k+1} \lambda_{k+1,j} x^j (1+x)^{n-j} (n)_j \end{aligned} \quad (5)$$

Note that the (5) shows that $\sum_{i=0}^n i^{k+1} \binom{n}{i} x^i = \sum_{j=1}^{k+1} \lambda_{k+1,j} x^j (1+x)^{n-j} (n)_j$ where $\lambda_{k+1,j}$ follows the recurrence relation

$$\lambda_{k+1,j} = j \lambda_{k,j} + \lambda_{k,j-1}.$$

As given in [23], Stirling numbers of second kind follow this recurrence.

$$\lambda_{k,j} = \frac{1}{j!} \sum_{i=0}^j j^i \binom{i}{j} (-1)^j \quad (6)$$

To get the k^{th} moment, we simply substitute $x = \frac{p}{1-p}$ in (4) and multiply by $(1-p)^n$. Hence we have the following theorem.

Theorem 4.

$$\mathbf{E}[X^k] = \sum_{j=1}^k \lambda_{k,j} p^j (n)_j$$

where $\lambda_{k,j}$ are as given in (6).

4 Estimators for counting k -cliques and k -clique covers in random graphs

In this section, we formally describe our estimators. The estimator for counting cliques is given in Algorithm 1. Note that it embeds the clique $\{v_1, \dots, v_k\}$ and outputs the inverse of probability of embedding it in this way into G . The estimator embeds one vertex at a time until the whole clique is embedded. If the algorithm gets stuck, it outputs 0. This process can be viewed as decomposing the clique into subgraphs C_1, C_2, \dots, C_k , where each C_i is the subgraph induced by the i^{th} numbered vertex v_i and its lower numbered neighbors. It is denoted by v_i .

We denote our randomized estimator by \mathcal{A} and let X be the output estimate. To get an fpras, we need that $\mathbf{E}_{\mathcal{A}}[X^2]/(\mathbf{E}_{\mathcal{A}}[X])^2$, also called the *critical ratio*, is polynomially bounded. We will bound a related quantity called *critical ratio of averages* given by $\text{Cr}(X) = \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X^2]]/(\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X]])^2$. Here, the outer expectation is over the graphs of $\mathcal{G}(n, p)$ and the inner expectation is over the coin tosses of the estimator. Our focus in this work will be to get a bound on critical ratio of averages. As shown in Prop. 1, this will also give a polynomial bound on the critical ratio itself. The proof of Prop. 1 follows from Corollary 2 of Theorem 5 from [30].

Consider any induced subgraph H_v of H with v vertices. Let $e_H(v) = \max_{H_v \subseteq H} \{|E(H_v)|\}$ of edge For stating the results, we need to define the following ratio for the template graph H .

$$\gamma = \gamma(H) = \max_{3 \leq v \leq n} \{e_H(v)/(v-2)\}.$$

Note that γ is closely related to the largest possible average degree of an induced subgraph of H . In our case, this is $(1+o(1)) \log n$ for the case of counting

cliques and $O(1)$ for counting clique covers. Let $C = C_H(G)$ denote the number of copies of H in G .

Theorem 5 ([30]). *Let H be a graph on n vertices and γ be as defined above. Let p be a constant. Suppose that the following conditions hold: $p \cdot \binom{n}{2} \rightarrow \infty$, $\sqrt{n}(1-p) \rightarrow \infty$ and $np^\gamma/\Delta^4 \rightarrow \infty$. Then, with high probability, a random graph $G \in \mathbb{G}(n, p \cdot \binom{n}{2})$ has a spanning subgraph isomorphic to H . In general, $C = C_H(G)$ satisfies*

$$\frac{\mathbf{E}[C^2]}{\mathbf{E}[C]^2} = 1 + o(1).$$

Remarks. Note that Theorem 5 holds for the spanning subgraphs of the random graphs. This assumption can easily be incorporated while embedding a single clique at any step. While embedding each clique, H is considered to be the n vertex graph which is the disjoint union of a clique and the isolated vertices in both the cases. Also, note that $np^\gamma/\Delta^4 \rightarrow \infty$ since γ and Δ are both bounded by $(1 + o(1)) \log n$. Therefore, all conditions of Theorem 5 are satisfied in our case. So we get the following corollary in our case.

Corollary 1. *Let $G \in \mathbb{G}(n, \Omega(n^2))$ and H be one of the following graphs*

(a) a clique of size $(1 + o(1)) \log_{\frac{1}{p}} n$ or (b) a cover of cliques of constant size,

Then $\frac{\mathbf{E}[C^2]}{\mathbf{E}[C]^2} = 1 + o(1)$, where C denotes the number of copies of H in G .

From the asymptotic equivalence between $\mathcal{G}(n, p)$ and $\mathbb{G}(n, m)$ (see e.g. [16, 28]), we have the following corollary.

Corollary 2. *Let $G \in \mathcal{G}(n, p)$ and H be one of the following graphs*

(a) a clique of size $(1 + o(1)) \log_{\frac{1}{p}} n$ or (b) a cover of cliques of constant size,

Then $C \geq \mathbf{E}[C]/\omega$, where $\omega = \omega(n)$ be a real valued function that goes to ∞ as $n \rightarrow \infty$.

Theorem 5 along with Corollary 1 and Corollary 2 yield the following proposition. The proof is identical to the one given for a similar proposition in [12], but we give it make the write-up self contained.

Proposition 1. *Let $G \in \mathcal{G}(n, p)$ and H be one of the following graphs*

(a) a clique of size $(1 + o(1)) \log_{\frac{1}{p}} n$ or (b) a cover of cliques of constant size.

Let X be the output of Algorithm Embeddings, and let p be a constant. Then, for a random graph $G \in \mathcal{G}(n, p)$ the critical ratio satisfies $\frac{\mathbf{E}[X^2]}{(\mathbf{E}[X])^2} \leq \omega^3 \frac{\mathbf{E}_G[\mathbf{E}_{\mathcal{A}}[X^2]]}{(\mathbf{E}_G[\mathbf{E}_{\mathcal{A}}[X]])^2}$, where $\omega = \omega(n)$ such that $\omega \rightarrow \infty$ as $n \rightarrow \infty$.

Proof. For the unbiased estimator \mathcal{A} , we have $C = \mathbf{E}_{\mathcal{A}}[X]$. Therefore, from Corollary 2, we have that $C = \mathbf{E}_{\mathcal{A}}[X] \leq \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X]]/\omega$ with high probability. Also, from Markov's inequality we have $\Pr[\mathbf{E}_{\mathcal{A}}[X^2] > \omega \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X^2]]] \leq 1/\omega$. Therefore with probability at least $1 - 1/\omega$, we have $\mathbf{E}_{\mathcal{A}}[X^2] \leq \omega \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X^2]]$. Our result follows from these inequalities.

In the rest of the paper, we focus on bounding the critical ratio of averages. The estimator for counting k -cliques is given in Algorithm 1. It embeds one clique of size $k = (1 + o(1)) \log_{\frac{1}{p}} n$ in G and outputs the inverse of probability of embedding. This is done by the procedure EMBED-CLIQUE, which is called only once in this case.

Algorithm 1 Count-cliques(G, k)

```

1: procedure EMBED-CLIQUE( $G, k$ )
2:    $i \leftarrow 0$  ▷  $i$  denotes the number of nodes already embedded in  $G$ 
3:    $v_1 \leftarrow \text{ArbitraryNode}(G)$  ▷ Arbitrarily assign a node from  $G$  to  $v_0$ 
4:   while  $i < k$  do
5:      $\mathcal{N}_i \leftarrow \text{CommonNeighbors}(\{v_1, \dots, v_i\})$ 
6:     if  $\mathcal{N}_i = \emptyset$  then
7:        $X \leftarrow 0$  ▷ Embedding algorithm has failed; so terminate
8:     end if
9:      $X_i \leftarrow |\mathcal{N}_i|$ 
10:     $v_{i+1} \leftarrow \text{RandomNode}(\mathcal{N}_i)$  ▷ uniformly randomly assign a node from  $\mathcal{N}_i$  to
11:     $X \leftarrow X \cdot X_i$ 
12:     $i \leftarrow i + 1$ 
13:  end while
14:  return  $X/(k!)$  ▷ Estimator outputs unbiased estimate of number of  $k$ -Cliques
15: end procedure

```

The estimator for counting k -clique covers of G is given in Algorithm 2. It uses the procedure EMBED-CLIQUE described in Algorithm 1 to embed each k -clique in the cover. This process is sequentially repeated until all the vertices are covered. In the end, it returns the inverse of probability of finding the cover, if successful. Note that this is the product of the probabilities of embedding the individual cliques in the cover.

5 Analysis of estimator for counting cliques and clique-covers in random graphs

In this section, we show a polynomial bound on the critical ratio of averages for the estimators in Algorithm 1 and Algorithm 2. Note that from Prop. 1, this is sufficient to bound the critical ratio of the estimator and hence get an fpras for counting k -cliques (for $k = (1 + o(1)) \log n$) and k -clique covers (for $k = O(1)$) in random graphs.

Algorithm 2 Count-clique-covers(G, k)

```

1:  $G_{res} \leftarrow G$ 
2:  $a \leftarrow (k!)^{\frac{n}{k}} \cdot (\frac{n}{k})!$  ▷ Size of the automorphism group of  $k$ -clique cover
3:  $X \leftarrow 1$ 
4: while  $G_{res} \neq \emptyset$  do
5:    $X \leftarrow X \cdot \text{EMBED-CLIQUE}(G_{res}, k)$ 
6:   if  $\text{EMBED-CLIQUE}(G_{res}, k) = 0$  then
7:      $X \leftarrow 0$  ▷ Embedding algorithm has failed; so terminate
8:   end if
9:    $G_{res} \leftarrow G \setminus \{v_1, \dots, v_k\}$  ▷ Remove the currently embedded clique
    $\{v_1, v_2, \dots, v_k\}$  from  $G$  to get  $G_{res}$ 
10: end while
11: return  $X/a$ 

```

5.1 Counting Cliques

In this section, we prove Theorem 1. In this case, the estimator embeds a single clique onto the base graph and outputs the inverse of probability of embedding the same. Let X , the random variable denoting the count, be the output of the estimator. The estimator selects first vertex in the graph arbitrarily and embeds one edge at a time until the whole clique is embedded. It outputs the inverse of probability of embedding if it goes through, else it outputs 0.

Let X_j corresponds to the number of ways to embed vertex j in the residual graph. Note that $X = X_1 \cdot X_2 \cdots X_k$. Now consider the term $\text{Cr}(X) = \mathbf{E}_G[\mathbf{E}_A[X^2]/\mathbf{E}_G[\mathbf{E}_A[X]]^2]$.

To estimate the critical ratio of averages, we need the definition of k -nesting, denoted by $N(k, n, p)$, as follows.

Definition 1 (*k -nesting*). A k -nesting is a function $N(k, n, p)$ that can be evaluated in the following recursive way.

(i) The 2-nesting is defined as

$$N(2, n, p) = n^2 \left(\sum_{i=1}^{n-1} i^2 \binom{n-1}{i} p^i (1-p)^{n-1-i} \right)$$

(ii) The k -nesting is defined as

$$N(k, n, p) = n^2 \left(\sum_{i=k-1}^{n-1} N(k-1, i, p) \binom{n-1}{i} p^i (1-p)^{n-1-i} \right)$$

Note that the embedding of a k -clique can be thought of as embedding i^{th} -vertex to get an i -clique from $i-1$ -clique for each $i \in \{1, 2, \dots, k\}$. So, we have the following observation.

Observation 1.

$$\mathbf{E}_G[\mathbf{E}_A[X_1^2 X_2^2 \cdots X_k^2]] = N(k, n, p) \quad (7)$$

Lemma 2 shows the exact structure of $N(k, \ell, p)$, which we use in getting the bound on the critical ratio.

Lemma 2.

$$N(k, \ell, p) = \sum_{j=k}^{2k-1} \ell(\ell)_j f_{k,j}(p)$$

Here $f_{k,j}(p)$ is a function in k, j, p that is independent of ℓ with the following properties.

- (i) $f_{k,k-i}(p) = 0$ for all $i \in \{1, \dots, k\}$ and $f_{k,2k+i}(p) = 0$ for all $i \geq 0$.
- (ii) $f_{k+1,j}(p) = p^{j-1}((j-1)f_{k,j-1}(p) + f_{k,j-2}(p))$.

Proof. We prove this by induction on k . For the base case, *i.e.* for $k = 2$ this is

$$\begin{aligned} N(2, \ell, p) &= \ell^2 \left(\sum_{i=1}^{\ell-1} i^2 \binom{\ell-1}{i} p^i (1-p)^{\ell-1-i} \right) \\ &= \ell^2 (\ell-1)_1 p + \ell^2 (\ell-1)_2 p^2 \text{ (using (4) with } k=2) \\ &= \ell(\ell)_2 p^{\binom{2}{2}} + \ell(\ell)_3 p^{2\binom{2}{2}} \end{aligned}$$

Suppose the claim is true for $N(i, \ell, p)$ for $i = \{1, 2, \dots, k\}$. We will show that the claim is true for $i = k+1$. From Definition 1 we have

$$\begin{aligned} N(k+1, \ell, p) &= \ell^2 \sum_{m=k}^{\ell-1} N(k, m, p) \binom{\ell-1}{m} p^m (1-p)^{\ell-1-m} \\ &= \ell^2 \sum_{m=k}^{\ell-1} \sum_{j=k}^{2k-1} (m(m)_j f_{k,j}(p)) \binom{\ell-1}{m} p^m (1-p)^{\ell-1-m} \\ &= \sum_{j=k}^{2k-1} \sum_{m=k}^{\ell-1} \left(\ell^2 (m)_j \binom{\ell-1}{m} p^m (1-p)^{\ell-1-m} \right) f_{k,j}(p) \text{ (interchanging the summations)} \\ &= \sum_{j=k}^{2k-1} (j \cdot \ell^2 (\ell-1)_j p^j + \ell^2 (\ell-1)_{j+1} p^{j+1}) f_{k,j}(p) \text{ (from Lemma 3, Eqn.(10))} \\ &= \sum_{i=k+1}^{2k+1} (j \cdot \ell(\ell)_i p^{i-1} + \ell(\ell)_{i+1} p^i) f_{k,i-1}(p) \text{ (using } \ell(\ell-1)_i = (\ell)_{i+1}), j+1=i, f_{k,2k}(p)=0) \\ &= \sum_{i=k+1}^{2k+1} p^{i-1} ((i-1)f_{k,i-1}(p) + f_{k,i-2}(p)) \ell(\ell)_i \text{ (rearranging the terms and using } f_{k,k-1}(p)=0) \\ &= \sum_{i=k+1}^{2k+1} f_{k+1,i}(p) \ell(\ell)_i \text{ (rearranging the terms and using } f_{k,k-1}(p)=0) \end{aligned} \tag{8}$$

The following lemma is used in the proof of Lemma 2.

Lemma 3.

$$\sum_{m=j}^n m(m)_j \binom{n}{m} x^{m-j} = j(n)_j (1+x)^{n-j} + (n)_{j+1} x (1+x)^{n-j-1} \quad (9)$$

In particular, if we multiply (9) by $x^j(1-p)^n$ and substitute $x = p/(1-p)$ we get

$$\sum_{m=j}^n m(m)_j \binom{n}{m} p^m (1-p)^{n-m} = j(n)_j p^j + (n)_{j+1} p^{j+1} \quad (10)$$

Proof. We prove the identity in (9) using induction. For $j = 0$ (base case) we need to show that $\sum_{m=0}^n m \binom{n}{m} x^m = nx(1+x)^{n-1}$, which holds from (2). For hypothesis, assume that (9) holds for j . We prove that it also holds for $j+1$ as follows. Differentiating (9) w.r.t. x gives

$$\begin{aligned} \sum_{m=j+1}^n m(m-j)(m)_j x^{m-j-1} &= j(n-j)(n)_j (1+x)^{n-j-1} (n)_{j+1} (1+x)^{n-j-1} \\ &\quad + (n-j-1)(n)_{j+1} x (1+x)^{n-j-2} \\ \sum_{m=j+1}^n m(m)_{j+1} x^{m-(j+1)} &= j(n)_{j+1} (1+x)^{n-j-1} + (n)_{j+1} (1+x)^{n-j-1} \\ &\quad + (n)_{j+2} x (1+x)^{n-j-2} \text{ (using } n(n-1)_i = (n)_{i+1}) \\ \sum_{m=j+1}^n m(m)_{j+1} x^{m-(j+1)} &= (j+1)(n)_{j+1} (1+x)^{n-j-1} + (n)_{j+2} x (1+x)^{n-j-2} \end{aligned}$$

Hence the identity holds for $j+1$.

The following lemma upper bounds $f_{k,k+i}(p)$ for $0 \leq i \leq k-1$

Lemma 4. For $k \geq 2$ $f_{k,2k-i-1}(p) \leq k^{2i} p^{\binom{k}{2} + \binom{k-i}{2}}$ where $0 \leq i \leq k-1$.

Proof. We will prove this claim using induction on k . Consider $k = 2$ for the base case. From Definition 1, we have $N(2, n, p) = n(n)_2 p + n(n)_3 p^2$. So, the claim holds. Now assume that the claim holds for all clique sizes up to $k-1$. Now, from (8), we have the following recurrence relation.

$$f_{k,i}(p) = p^{i-1}((i-1)f_{k-1,i-1}(p) + f_{k-1,i-2}(p)) \quad (11)$$

First we prove for $i \geq 1$. Using (11), we have

$$\begin{aligned} f_{k,2k-i-1}(p) &= p^{2k-i-2}((2(k-1)-i)f_{k-1,2(k-1)-(i-1)-1}(p) + f_{k-1,2(k-1)-i-1}(p)) \\ &\leq p^{2k-i-2} \left((2(k-1)-i)(k-1)^{2(i-1)} p^{\binom{k-1}{2} + \binom{k-i}{2}} + (k-1)^{2i} p^{\binom{k-1}{2} + \binom{k-i-1}{2}} \right) \\ &= (k-1)^{2i} p^{\binom{k}{2} + \binom{k-1}{2}} \left(1 + p^{k-i-1} \left(\frac{2}{k-1} - \frac{i}{(k-1)^2} \right) \right) \\ &\leq (k-1)^{2i} p^{\binom{k}{2} + \binom{k-1}{2}} \left(1 + \frac{2}{k-1} \right) \\ &= ((k-1)^{2i} + 2(k-1)^{2i-1}) p^{\binom{k}{2} + \binom{k-1}{2}} \leq k^{2i} p^{\binom{k}{2} + \binom{k-1}{2}} \text{ (for } i \geq 1) \end{aligned}$$

Now we show that $f_{k,2k-1} = p^{2\binom{k}{2}}$. From (11), we have $f_{k,2k-1}(p) = p^{2(k-1)}((2k-2)f_{k-1,2k-2}(p) + f_{k-1,2k-3}(p)) = p^{2(k-1)}f_{k-1,2k-3}(p)$ since $f_{k-1,2k-2}(p) = 0$. Applying the recurrence repeatedly, we get the desired relation.

Now we bound $\text{Cr}(X)$ which is the same as $\frac{N(k,n,p)}{\left((n)_k p^{\binom{k}{2}}\right)^2}$. We have

$$\text{Cr}(X) = \frac{N(k,n,p)}{\left((n)_k p^{\binom{k}{2}}\right)^2} = \sum_{j=k}^{2k-1} \frac{n(n)_i f_{k,j}(p)}{\left((n)_k p^{\binom{k}{2}}\right)^2} = \sum_{i=0}^{k-1} \frac{n(n)_{2k-i-1} f_{k,2k-i-1}(p)}{\left((n)_k p^{\binom{k}{2}}\right)^2} \quad (12)$$

Lemma 5 immediately proves Theorem 1.

Lemma 5. For $k = (1 + o(1)) \log_{\frac{1}{p}} n$, $\text{Cr}(X) = \sum_{i=0}^{k-1} \frac{n(n)_{2k-i-1} f_{k,2k-i-1}(p)}{\left((n)_k p^{\binom{k}{2}}\right)^2}$ is upper bounded by $\text{poly}(n)$.

Proof. Consider the ratio $\frac{\ell(\ell)_{2k-i-1} f_{k,2k-i-1}(p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2}$ for a fixed i . Here we have $\ell = n$.

As we shall see, ℓ changes for the k -Clique cover. For $i = 0$, this is $\frac{\ell(\ell)_{2k-1}}{\left((\ell)_k\right)^2}$ since $f_{k,2k-1} = p^{2\binom{k}{2}}$. Note that $\frac{\ell(\ell)_{2k-1}}{\left((\ell)_k\right)^2} \leq 1$. Now we consider $i \geq 1$.

$$\begin{aligned} \frac{\ell(\ell)_{2k-i-1} f_{k,2k-i-1}(p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2} &= \left(\prod_{j=1}^{k-i-1} \frac{(\ell - (k-1) - j)}{(\ell - j)} \right) \left(\frac{f_{k,2k-i-1}(p)}{\prod_{r=1}^i (\ell - k + r)} \right) \frac{1}{p^{2\binom{k}{2}}} \\ &\leq \left(\frac{\ell - k}{\ell - 1} \right)^{k-i-1} \left(\frac{k^{2i} p^{\binom{k}{2} + \binom{k-i}{2}}}{(\ell - k + 1)^i} \right) \frac{1}{p^{2\binom{k}{2}}} \\ &= \left(\frac{\ell - k}{\ell - 1} \right)^{k-i-1} \left(\frac{k^2}{\ell - k + 1} \right)^i \frac{1}{p^{\binom{k}{2} - \binom{k-i}{2}}} \\ &= \left(\frac{\ell - k}{\ell - 1} \right)^{k-i-1} \left(\frac{k^2}{\ell - k + 1} \left(\frac{1}{p} \right)^{k - \left(\frac{i+1}{2}\right)} \right)^i = h(i) \quad (13) \end{aligned}$$

The first inequality above uses Lemma 4, $\frac{\ell-k}{\ell-1} \geq \frac{\ell-k-j}{\ell-1-j}$ and $\ell - k + j \geq \ell - k + 1$ for all $1 \leq j \leq k-1$. Note that $\left(\frac{\ell-k}{\ell-1} \right)^{k-i-1} \leq 1$. So, we have $h(i) \leq \left(\frac{k^2}{\ell-k+1} \left(\frac{1}{p} \right)^{k - \left(\frac{i+1}{2}\right)} \right)^i$, where $h(i)$ is as defined in (13). Note that for $i = (1 + o(1)) \log n$, $\left(\frac{k^2}{\ell-k+1} \left(\frac{1}{p} \right)^{k - \left(\frac{i+1}{2}\right)} \right)^i$ is polynomially bounded for all $0 \leq i \leq k-1$. Therefore $\text{Cr}(X)$ is polynomially bounded.

Lemma 6. For $k = (1 + o(1)) \log_{\frac{1}{p}} n$, $h(i) = \left(\frac{k^2}{\ell-k+1} \left(\frac{1}{p} \right)^{k - \left(\frac{i+1}{2}\right)} \right)^i$ is polynomially bounded for all $0 \leq i \leq k-1$.

Proof. First note that

$$\left(\frac{k^2}{n-k+1} \left(\frac{1}{p} \right)^{k - \left(\frac{i+1}{2} \right)} \right)^i = \left(\frac{1}{p} \right)^{\left(2i \log_{\frac{1}{p}} k - i \log_{\frac{1}{p}} (n-k+1) + ki - \frac{i(i+1)}{2} \right)}$$

Let $g(i) = 2i \log_{\frac{1}{p}} k - i \log_{\frac{1}{p}} (n-k+1) + ki - \frac{i(i+1)}{2}$, where $k = (1 + \varepsilon_n) \log_{\frac{1}{p}} n$. This function is maximized at the point where $\partial g(i)/\partial i = 0$, which happens at $i \approx 2 \log_{\frac{1}{p}} \log_{\frac{1}{p}} n + \varepsilon_n \log_{\frac{1}{p}} n$. At this point, $g(i) \approx 2 \left(\log_{\frac{1}{p}} \log_{\frac{1}{p}} n \right)^2 + \frac{\varepsilon_n^2}{2} (\log_{\frac{1}{p}} n)^2$. Note that $h(i) = \left(\frac{1}{p} \right)^{g(i)}$ is polynomially bounded only when $\varepsilon_n = O\left(\frac{1}{\sqrt{\log_{\frac{1}{p}} n}}\right)$.

5.2 Clique cover counting

As noted earlier in Prop. 1, we focus on bounding the *critical ratio of averages* given by $\text{Cr}(X) = \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X^2]] / (\mathbf{E}_{\mathcal{G}}[E_{\mathcal{A}}[X]])^2$ for Algorithm 1.

The estimator embeds one clique at a time, by selecting a vertex at random at first and then embedding each edge till k vertices of the clique are embedded. A crucial observation is that the residual graph, after embedding a clique still remains random with edge probability p . Finally, the estimator sequentially embeds n/k cliques to get the clique cover and outputs the inverse of probability of getting this clique cover, if the embedding procedure goes through, otherwise it outputs 0. Note that this is the product of the inverse of the probabilities for embedding each clique. Let K_i denote the random variable corresponding to the estimate of the number of embeddings of the i^{th} clique in the residual graph, which is a random graph from $\mathcal{G}(n - ki - k, p)$. Note that K_i is independent from K_j for $i \neq j$ and $X = K_1 \cdot K_2 \cdots K_{\frac{n}{k}}$. Therefore we have the following equation.

$$(\mathbf{E}_{\mathcal{G}}[E_{\mathcal{A}}[X]])^2 = \prod_{i=1}^{\frac{n}{k}} (E[K_i])^2 \quad (14)$$

Note that the equality follows from the fact that after embedding each k -clique, the residual graph still remains random with edge probability p . Now, we bound the numerator, *i.e.*, $\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X^2]]$.

$$\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X^2]] = \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[K_1^2 K_2^2 \cdots K_{\frac{n}{k}}^2]] = \mathbf{E}_{\mathcal{G}}[E_{\mathcal{A}}[K_1^2]] \cdot \mathbf{E}_{\mathcal{G}}[E_{\mathcal{A}}[K_2^2]] \cdots \mathbf{E}_{\mathcal{G}}[E_{\mathcal{A}}[K_{\frac{n}{k}}^2]]$$

Let X_j corresponds to the number of ways to embed vertex j in the residual graph. Note that $K_i = X_{ki-k+1} \cdot X_{ki-k+2} \cdots X_{ki}$.

Now consider the term $\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[K_i^2]] = \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X_{k(i-1)+1}^2 X_{k(i-1)+2}^2 \cdots X_{ki}^2]]$. Note that in this case, we have

$$\text{Cr}(K_i) = \frac{\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X_{ki-k+1}^2 X_{ki-k+2}^2 \cdots X_{ki}^2]]}{\mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[X_{ki-k+1} X_{ki-k+2} \cdots X_{ki-k+k}]]^2} = \frac{N(k, n - ki + k, p)}{\left((n - ki)_k p^{\binom{k}{2}} \right)^2} \quad (16)$$

We show in Lemma 7 that $\frac{N(k, \ell, p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2}$ is bounded by $1 + O\left(\frac{1}{n - ki + 1}\right)$ for all $i \in \{1, 2, \dots, \frac{n}{k}\}$, where $\ell = n - ki + k$.

Lemma 7. *For large ℓ , constant k and constant p we have*

$$\text{Cr}(K_i) = \sum_{j=0}^{k-1} \frac{\ell(\ell)_{2k-j-1} f_{k, 2k-j-1}(p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2} \leq 1 + O\left(\frac{1}{\ell - k + 1}\right)$$

Proof. Consider the ratio $\frac{\ell(\ell)_{2k-j-1} f_{k, 2k-j-1}(p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2}$ for a fixed j . For $j = 0$, this is $\frac{\ell(\ell)_{2k-1}}{\left((\ell)_k\right)^2}$ since $f_{k, 2k-1} = p^{2\binom{k}{2}}$. Note that $\frac{\ell(\ell)_{2k-1}}{\left((\ell)_k\right)^2} \leq 1$. Now we consider $j \geq 1$. As shown in (13), we have

$$\frac{\ell(\ell)_{2k-1-j} f_{k, 2k-j-1}(p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2} \leq h(j) = \left(\frac{\ell - k}{\ell - 1}\right)^{k-j-1} \left(\frac{k^2}{\ell - k + 1} \left(\frac{1}{p}\right)^{k - \left(\frac{j+1}{2}\right)}\right)^j$$

To prove the lemma, we handle the cases of $j \leq 2$ and $j \geq 2$ separately. First we handle the latter case. For $j \geq 2$, we prove that $h(j) \leq \frac{1}{(k-2)(\ell-k+1)}$. In other words, we prove that $\log h(j) + \log(k-2) + \log(\ell-k+1) < 0$ for constant k .

Let $y(j) = \log(h(j)) = (k-1-j)(\log(\ell-k) - \log(\ell-1)) + j(2\log k - \log(\ell-k+1)) + j\left(k - \frac{j+1}{2}\right) \log \frac{1}{p}$. Consider the continuous function $y(x) = (k-1-x)(\log(\ell-k) - \log(\ell-1)) + x(2\log k - \log(\ell-k+1)) + x\left(k - \frac{x+1}{2}\right) \log \frac{1}{p}$. Therefore we have

$$\begin{aligned} y'(x) &= \frac{\partial y(x)}{\partial x} = \log(\ell-1) - (\log(\ell-k) + \log(\ell-k+1)) + 2\log k + \left(k - x + \frac{1}{2}\right) \log \frac{1}{p} \\ &\leq \log(\ell-1) - 2(\log(\ell-k)) + 2\log k + \left(k - x + \frac{1}{2}\right) \log \frac{1}{p} \end{aligned}$$

Observe that for large ℓ and for constant k , the term $-2(\log(\ell-k))$ dominates all the other terms, so $y'(x) < 0$ for $1 \leq x \leq k-1$. Therefore $y(x)$ is a decreasing function. We analyze cases $j = 1$ and $j \geq 2$ separately. First we analyze latter case. We prove that $y(2) \leq \log\left(\frac{1}{(k-2)(\ell-k+1)}\right)$, which implies that $y(j) = \log\left(\frac{1}{(k-2)(\ell-k+1)}\right)$ for $2 \leq j \leq k-1$. This proves that $h(i) = \frac{1}{(k-2)(\ell-k+1)}$ for $j \geq 2$, eventually proving that

$$\sum_{j=2}^{k-1} \frac{\ell(\ell)_{2k-j-1} f_{k, 2k-j-1}(p)}{\left((\ell)_k p^{\binom{k}{2}}\right)^2} \leq \frac{1}{\ell - k + 1} \quad (17)$$

Consider the function $g(\ell) = y(2) + \log(k-2) + \log(\ell-k+1)$. So we have

$$\begin{aligned} g(\ell) &= (k-3)(\log(\ell-k) - \log(\ell-1)) + 2(2\log k - \log(\ell-k+1)) \\ &\quad + (2k-3)\log \frac{1}{p} + \log(\ell-k+1) + \log(k-2) \\ &\leq 4\log k + (2k-3)\log \frac{1}{p} + \log(k-2) - \log(\ell-k+1) \end{aligned}$$

Note that for constant k , this is smaller than 0 for large enough ℓ . Therefore $g(\ell) < 0$, hence the claim.

Now we do the analysis for $j = 1$. We calculate $f_{k,2k-2}(p)$ using the recurrence.

$$\begin{aligned} f_{k,2k-2}(p) &= p^{2k-3} ((2k-3)f_{k-1,2(k-1)-1}(p) + f_{k-1,2(k-1)-2}(p)) \\ &= (2k-3)p^{2k-3+2\binom{k-1}{2}} + p^{2k-3}f_{k-1,2(k-1)-2}(p) \quad (\text{using } f_{k-1,2(k-1)-1} = p^{2\binom{k-1}{2}}) \\ &= (2k-3)p^{2\binom{k}{2}-1} + p^{2k-3+2k-5} ((2k-5)f_{k-2,2(k-2)-1}(p) + f_{k-2,2(k-2)-2}(p)) \\ &= (2k-3)p^{2\binom{k}{2}-1} + (2k-5)p^{2\binom{k}{2}-2} + p^{2k-3+2k-5}f_{k-2,2(k-2)-2}(p) \end{aligned}$$

Going on as shown in the above equation, we get $f_{k,2k-2}(p) = \sum_{m=1}^{k-1} (2(k-m) - 1)p^{2\binom{k}{2}-m}$. Therefore we have

$$\frac{\ell(\ell)_{2k-2}}{((\ell)_k p^{\binom{k}{2}})^2} = \frac{\ell(\ell)_{2k-2}}{((\ell)_k)^2} \left(\sum_{m=1}^{k-1} \frac{2(k-m)-1}{p^m} \right)$$

Note that

$$\begin{aligned} \left(\sum_{m=1}^{k-1} \frac{2(k-m)-1}{p^m} \right) &= \frac{1}{\frac{1}{p}-1} \left(2 \left(\frac{\frac{1}{p^k}-1}{\frac{1}{p}-1} \right) + \frac{1}{p^k} - \left(\frac{2k+2p-1}{p} \right) \right) \\ &\leq \frac{C}{p^k} \quad (\text{for large enough constant } C) \end{aligned} \tag{18}$$

Note that for constant k , $\frac{C}{p^k} = C'$ is a constant. Therefore, using (17) and (18) we have

$$\sum_{j=0}^{k-1} \frac{\ell(\ell)_{2k-j-1} f_{k,2k-j-1}(p)}{((\ell)_k p^{\binom{k}{2}})^2} \leq 1 + \left(\frac{C' + 1}{\ell - k + 1} \right)$$

Hence the lemma.

Note that Lemma 7 shows that $\text{Cr}(K_i) = \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[K_i^2]] / \mathbf{E}_{\mathcal{G}}[\mathbf{E}_{\mathcal{A}}[K]]^2 = 1 + O\left(\frac{1}{n-ki+1}\right)$. Note that Theorem 3 follow from Lemma 7 since $\prod_{i=1}^{\frac{n}{k}} \text{Cr}(K_i) = \text{poly}(n)$ in this case.

6 Conclusion and open problems

In this work, we show the first fpras for counting k -cliques, where $k = (1 + o(1)) \log_{\frac{1}{p}} n$ and k -clique covers (for constant k) in random graphs, using the unbiased estimators that are very simple to describe. Both problems are $\#P$ -complete in general for the respective values of k . Getting a fpras for these problems over general graphs is a long standing open problem. Here are some specific open problems that we think are worth investigating.

1. The problem of counting clique is still open for counting cliques of size greater $(1 + o(1)) \log_{\frac{1}{p}} n$. Solving this will resolve the open problem of Frieze and McDiarmid ([10]) completely, though, this is probably very hard to solve [6].
2. Another specific problem to resolve here is to count clique covers of super-constant sized cliques.
3. The determinant based estimators usually have smaller worst case running times in fpras (e.g. [5]) for random graphs. It is unclear to us how to obtain any determinant based unbiased estimators for the clique and clique cover counting problems.

References

1. Tom A.B.Snijders, Philippa E. Pattison, , Garry L. Robins, and Mark S. Handcock. New specifications for exponential random graph models. *Sociological Methodology*, 36(1):99–153, 2006.
2. Ivona Bezáková, Daniel Stefankovic, Vijay V. Vazirani, and Eric Vigoda. Accelerating simulated annealing for the permanent and combinatorial counting problems. *SIAM J. Comput.*, 37(5):1429–1454, 2008.
3. Andrei Z. Broder. How hard is to marry at random? (on the approximation of the permanent). In *Proceedings of the 18th Annual ACM Symposium on Theory of Computing, May 28-30, 1986, Berkeley, California, USA*, pages 50–58, 1986.
4. Venkat Chandrasekaran, Misha Chertkov, David Gamarnik, Devavrat Shah, and Jinwoo Shin. Counting independent sets using the bethe approximation. *SIAM J. Discrete Math.*, 25(2):1012–1034, 2011.
5. Steve Chien. A determinant-based algorithm for counting perfect matchings in a general graph. In J. Ian Munro, editor, *Proceedings of the Fifteenth Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2004, New Orleans, Louisiana, USA, January 11-14, 2004*, pages 728–735. SIAM, 2004.
6. Amin Coja-Oghlan and Charilaos Efthymiou. On independent sets in random graphs. In Dana Randall, editor, *Proceedings of the Twenty-Second Annual ACM-SIAM Symposium on Discrete Algorithms, SODA 2011, San Francisco, California, USA, January 23-25, 2011*, pages 136–144. SIAM, 2011.
7. Gerard Cornuejols, David Hartvigsen, and William R. Pulleyblank. Packing subgraphs in a graph. In *Operations Research Letters*, volume 1, pages 139–143, 1982.
8. Martin E. Dyer, Alan M. Frieze, and Mark Jerrum. On counting independent sets in sparse graphs. *SIAM J. Comput.*, 31(5):1527–1541, 2002.
9. Paul Erdős and Alfréd Rényi. On the evolution of random graphs. *Publ. Math. Inst. Hung. Acad. Sci.*, 5:17–61, 1960.

10. Alan M. Frieze and Colin McDiarmid. Algorithmic theory of random graphs. *Random Struct. Algorithms*, 10(1-2):5–42, 1997.
11. Martin Fürer and Shiva Prasad Kasiviswanathan. Approximately counting perfect matchings in general graphs. In Camil Demetrescu, Robert Sedgewick, and Roberto Tamassia, editors, *Proceedings of the Seventh Workshop on Algorithm Engineering and Experiments and the Second Workshop on Analytic Algorithmics and Combinatorics, ALENEX / ANALCO 2005, Vancouver, BC, Canada, 22 January 2005*, pages 263–272. SIAM, 2005.
12. Martin Fürer and Shiva Prasad Kasiviswanathan. Approximately counting embeddings into random graphs. In Ashish Goel, Klaus Jansen, José D. P. Rolim, and Ronitt Rubinfeld, editors, *APPROX-RANDOM*, volume 5171 of *Lecture Notes in Computer Science*, pages 416–429. Springer, 2008.
13. Chris D. Godsil and Ivan Gutman. On the matching polynomial of a graph. pages 241–249, 1981.
14. G. R. Grimmett and C. J. H. McDiarmid. On colouring random graphs. 77:313–324, 1975.
15. Cornelis Hoede and Xueliang Li. Clique polynomials and independent set polynomials of graphs. *Discrete Mathematics*, 125(1-3):219–228, 1994.
16. Svante Janson, Tomasz Łuczak, and Andrzej Ruciński. *Random Graphs*. Wiley-Interscience, 2000.
17. Mark Jerrum and Alistair Sinclair. Approximating the permanent. *SIAM J. Comput.*, 18(6):1149–1178, 1989.
18. Mark Jerrum, Alistair Sinclair, and Eric Vigoda. A polynomial-time approximation algorithm for the permanent of a matrix with nonnegative entries. *J. ACM*, 51(4):671–697, 2004.
19. Anders Johansson, Jeff Kahn, and Van H. Vu. Factors in random graphs. *Random Struct. Algorithms*, 33(1):1–28, 2008.
20. Narendra Karmarkar, Richard M. Karp, Richard J. Lipton, László Lovász, and Michael Luby. A monte-carlo algorithm for estimating the permanent. *SIAM J. Comput.*, 22(2):284–293, 1993.
21. Richard M. Karp. Reducibility among combinatorial problems. In *Proceedings of a symposium on the Complexity of Computer Computations, held March 20-22, 1972, at the IBM Thomas J. Watson Research Center, Yorktown Heights, New York.*, pages 85–103, 1972.
22. P.W. Kasteleyn. The statistics of dimers on a lattice, i., the number of dimer arrangements on a quadratic lattice. *Physica*, 27:166–1672, 1961.
23. Andreas Knoblauch. Closed-form expressions for the moments of the binomial probability distribution. *SIAM Journal of Applied Mathematics*, 69(1):197–204, 2008.
24. Michihiro Kuramochi and George Karypis. Discovering frequent geometric subgraphs. *Inf. Syst.*, 32(8):1101–1120, 2007.
25. László Lovász and Michael D. Plummer. *Matching Theory*. MS Chelsea Publishing.
26. Michael Luby and Eric Vigoda. Approximately counting up to four (extended abstract). In *Proceedings of the Twenty-ninth Annual ACM Symposium on Theory of Computing, STOC '97*, pages 682–687, New York, NY, USA, 1997. ACM.
27. Natasa Przulj, Derek G. Corneil, and Igor Jurisica. Efficient estimation of graphlet frequency distributions in protein-protein interaction networks. *Bioinformatics*, 22(8):974–980, 2006.
28. Lars Eilstrup Rasmussen. Approximating the permanent: A simple approach. *Random Struct. Algorithms*, 5(2):349–362, 1994.

29. Lars Eilstrup Rasmussen. Approximately counting cliques. *Random Struct. Algorithms*, 11(4):395–411, 1997.
30. Oliver Riordan. Spanning subgraphs of random graphs. *Combinatorics, Probability & Computing*, 9(2):125–148, 2000.
31. Seinosuke Toda. PP is as hard as the polynomial-time hierarchy. *SIAM J. Comput.*, 20(5):865–877, 1991.
32. Julian R. Ullmann. An algorithm for subgraph isomorphism. *J. ACM*, 23(1):31–42, 1976.
33. Leslie G. Valiant. The complexity of computing the permanent. *Theor. Comput. Sci.*, 8:189–201, 1979.
34. Dror Weitz. Counting independent sets up to the tree threshold. In *Proceedings of the Thirty-eighth Annual ACM Symposium on Theory of Computing*, STOC '06, pages 140–149, New York, NY, USA, 2006. ACM.